

Proviso with Pseudowords: A Pilot Study

2020 Nov 16

Link:

https://spellout.net/ibexexps/sherrychen/proviso2020_sur/experiment.html

Participants

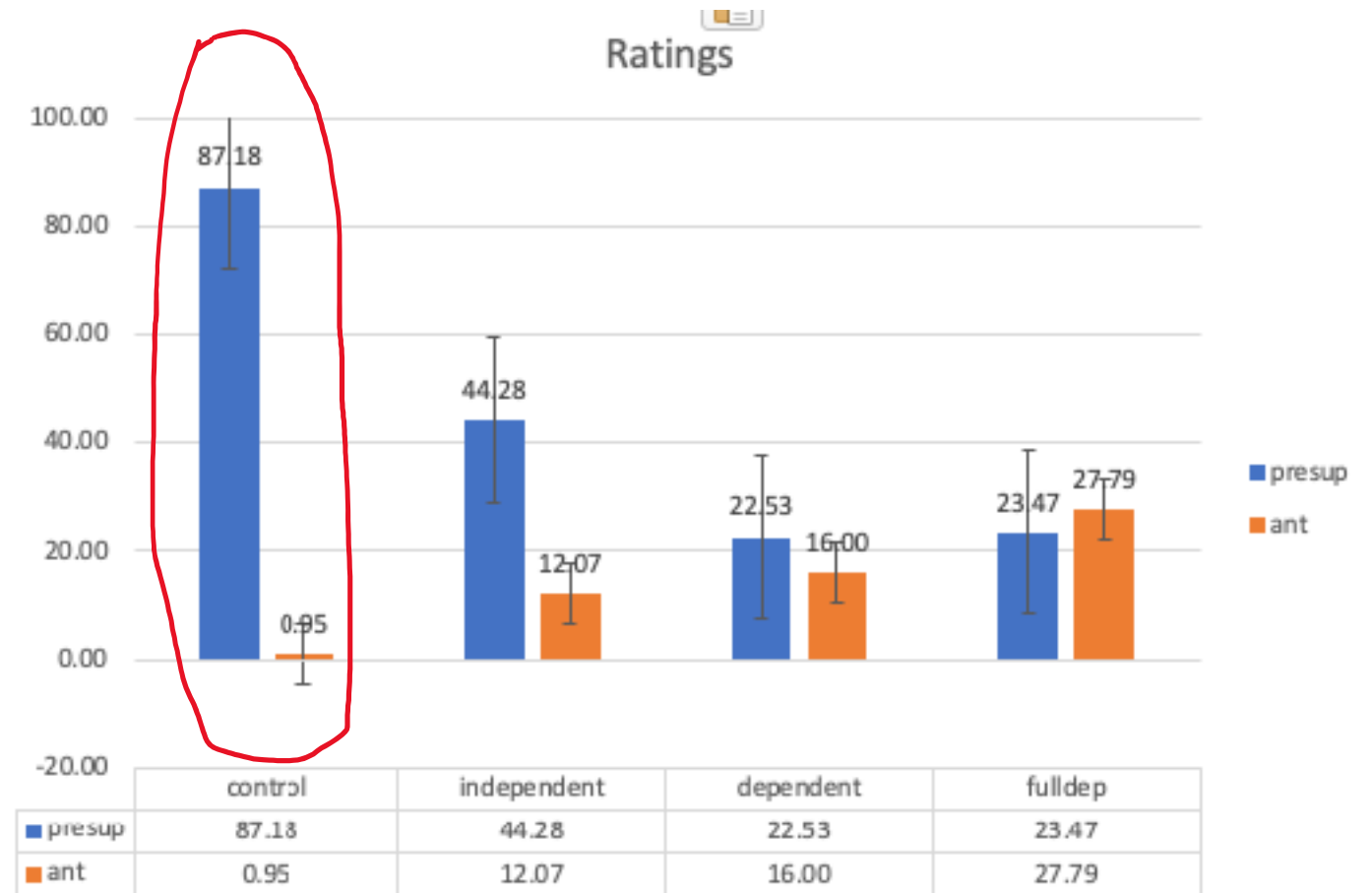
- 22 participants recruited from Twitter & Facebook
- Filtering based on the control item “John is a glorp and he has a dord”
 - “...John doesn’t have a dord”:
 - 14 got 80+ (11 got almost 100)
 - 5 got 60+
 - 3 got 0-25
 - “...John is a glorp”: 20 got 0, 1 got 5, one got 12
- **Instruction:** example said “**very** surprised” instead of “totally”!!!
 - Due to noise (handler sensitivity & some people not so strict), I’m only excluding those who responded below 60 for “...John doesn’t have a dord”.
 - This leaves us with **19** participants.

“Could be either way” & “surprised”?

- There seems to be different understandings of how “surprised” people are when things “could go either way”:
 - Some (like us) think it’s 50 (treating it like probability).
 - BUT many actually think it’s 0, because there’s no prior contradictory information and we are just learning something new, so nothing to be really surprised about; **for 50, you’ll be “*somewhat surprised*”, since 100 is “*totally surprised*”**
 - Comments: “Took me a while to catch on to *if* alien is an xyz she will xxx **but on reflection there should be no surprise that alien is type A or not type A. It's all about assumptions**”
- Questions:
 - Does it have to do with 0 being the initial state of the handler?
 - The example we gave in the instruction?
 - Will a different wording change it, e.g. surprised vs. unexpected?

Ratings: Control

- “John doesn’t have a dord” is close to 90, because it was a contradiction
- “John is a glorp” is close to 0, because it was consistent with the prior information



Ratings: Independent

- “John doesn’t have a dord” is around 44; people are somewhat surprised, but obviously not as surprised as seeing a contradiction;
- “John is a glorp” is around 12



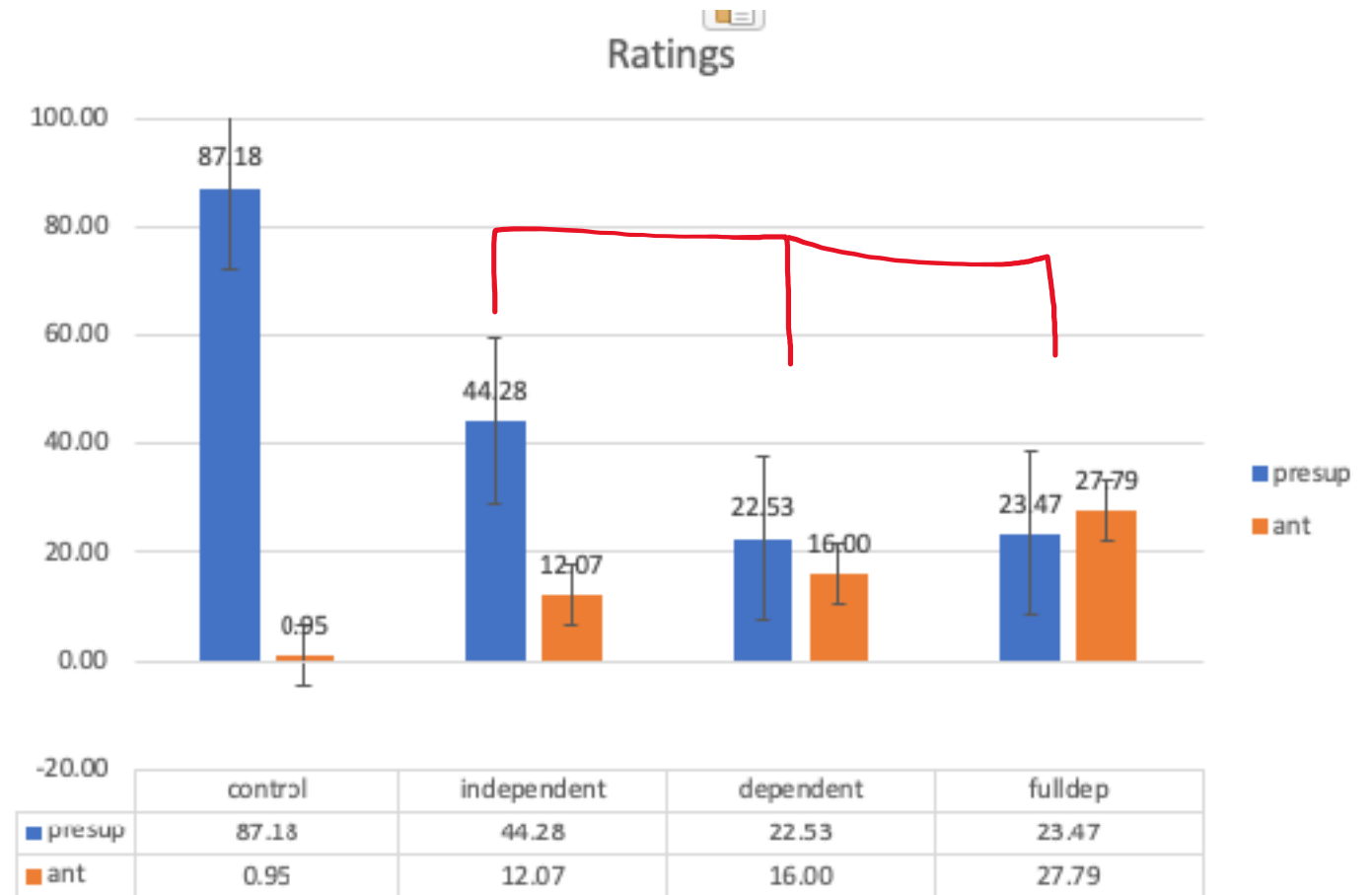
Ratings: Dependent & FullDep

- Dependent:
 - “John doesn’t have a dord” is 22.5;
 - “John is a glorp” is 16
- FullDep
 - “John doesn’t have a dord” is 23.5
 - “John is a glorp” is 28
- “*John doesn’t have a dord*”:
 - Quite similar
 - Both lower than Independent



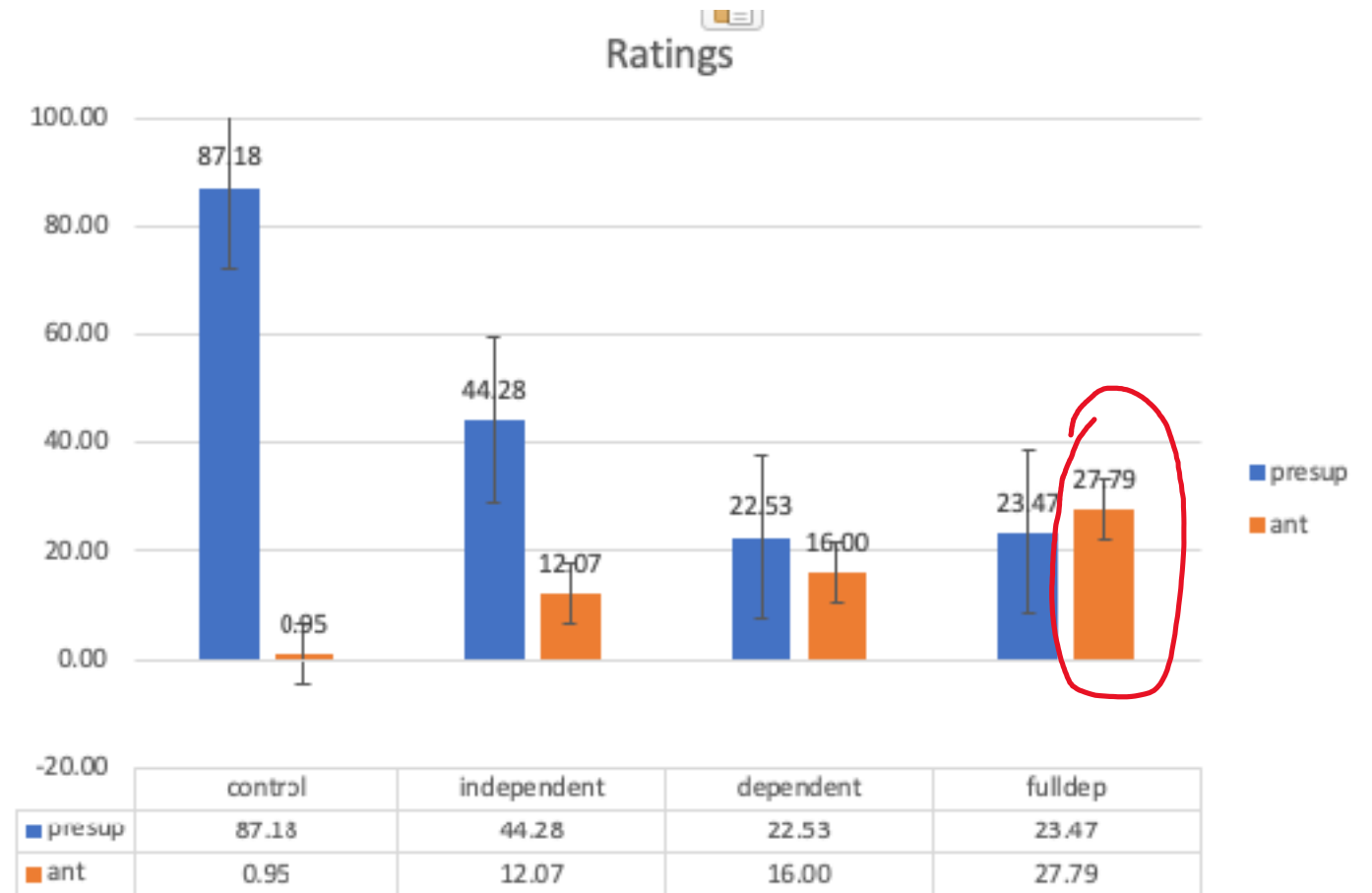
Differences between ratings for presup

- People seem to be **more surprised** at “John doesn’t have a dord” in **Independent**: might suggest that they strengthened to a non-conditional inference “John has a dord” in **Independent**, at least more so than they did in Dependent & FullDep
- Not a huge numerical difference here (around 22), but we only have 19 observations. Need more power to test for statistical significance!



Differences between ratings for Ant

- Tiny difference between Ant in Independent & Dependent: people are not surprised when they learn “John is a glorp”
- Ant in FullDep: ???
 - Driven by some **outliers**: 1 weird dude chose 80+, 2 chose 95+
 - This is not the same as Independent Presup, where you can observe a genuine different distribution!



Rating differentials

- Differentials of presup-ant is larger in Independent than Dependent & FullDep
- Differentials of presup-ant are similar in Dependent and FullDep
- But negative in FullDep, because of the weirdly highish ratings for Ant there.....
 - Maybe we don't have to worry about it so seriously because it's driven by outliers & we have little data
 - Item filtering? e.g. Ant has to be < 50 or something

